# EMPIRICAL BAYES METHODS FOR DYNAMIC FACTOR MODELS

S. J. Koopman and G. Mesters*

*Abstract*—We consider the dynamic factor model where the loading matrix, the dynamic factors, and the disturbances are treated as latent stochastic processes. We present empirical Bayes methods that enable the shrinkage-based estimation of the loadings and factors. We investigate the methods in a large Monte Carlo study where we evaluate the finite sample properties of the empirical Bayes methods for quadratic loss functions. Finally, we present and discuss the results of an empirical study concerning the forecasting of U.S. macroeconomic time series using our empirical Bayes methods.

## I. Introduction

CONSIDER the dynamic factor model for $N$ variables and time series length $T$ given by

$$y_{i,t} = \lambda_i' \alpha_t + \epsilon_{i,t}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \tag{1}$$

where $y_{i,t}$ is the observation corresponding to variable $i$ and time period $t$, $\lambda_i$ is the $r \times 1$ vector of factor loadings, $\alpha_t$ is the $r \times 1$ vector of dynamic factors, and $\epsilon_{i,t}$ is the disturbance term. The aim is to decompose the vector of time series observations $y_t = (y_{1,t}, \ldots, y_{N,t})'$ into two independent components: a common component that is driven by $r$ common dynamic processes in the vector $\alpha_t$ and an idiosyncratic component represented by the $N$ independent time series processes $\epsilon_{i,t}$. Dynamic factor models are typically used for macroeconomic forecasting or structural analysis. (See, e.g., Stock & Watson, 2002b, and Bernanke, Boivin, & Eliasz, 2005. The reviews of Bai & Ng, 2008, and Stock & Watson, 2011, provide more discussion and references.)

In this paper, we develop parametric empirical Bayes methods for the estimation of the loadings and the factors. We treat the loadings, factors, and disturbances as latent stochastic processes and estimate the hyperparameters that pertain to the distributions of the latent processes by maximum likelihood. The development of empirical Bayes methods for dynamic factor models is motivated by three related developments in the literature.

First, recent contributions by Doz, Giannone, and Reichlin (2012), Bai and Li (2012, 2016), and Banbura and Modugno (2014) have shown that the maximum likelihood estimates for the loadings and factors are more accurate when compared to the principal components estimates. The maximum likelihood method relies on the estimation of a large number of deterministic parameters (see Jungbacker & Koopman, 2015). While this estimation method is shown to be computationally feasible, estimation accuracy can potentially be improved by using shrinkage methods. Second, when $\alpha_t$ is observed, such that the model reduces to a multivariate regression model, James-Stein-type shrinkage estimators for $\lambda_i$ are known to outperform maximum likelihood estimators under various conditions for mean-squared error loss functions (see James & Stein, 1961; Efron & Morris, 1973; Knox, Stock, & Watson, 2004; Efron, 2010). We empirically investigate whether this remains the case when $\alpha_t$ is unobserved. Third, the recent study of Kim and Swanson (2014) shows that a dimension-reduction method based on a factor model specification, combined with shrinkage-based parameter estimation, leads to an empirically good method for forecasting macroeconomic and financial variables. We also combine dimension-reduction and shrinkage-based parameter estimation, but we consider a likelihood-based framework that is likely to improve on the principal components method (see Doz et al., 2012, and Bai & Li, 2012).

To facilitate the implementation of the empirical Bayes methods, we assume that the loading vectors are normally and independently distributed, while the dynamic factors are specified as a stationary vector autoregressive process. The stochastic assumptions for both the loadings and factors have been considered earlier in Bayesian dynamic factor analysis (see, e.g., Aguilar & West, 2000). However, they contrast with most other specifications, where the elements of the loading vectors and possibly the factors are treated as deterministic unknown variables (see Stock & Watson, 2011). For this model specification, we estimate the loadings and factors using filtering methods and the vector of unknown parameters, which is associated with the stochastic processes for $\lambda_i$, $\alpha_t$, and $\epsilon_{i,t}$, using the maximum likelihood method. The implementation of empirical Bayes methods for the dynamic factor model is nontrivial given the product of stochastic variables $\lambda_i$ and $\alpha_t$ in equation (1). Standard state-space methods, as discussed in Durbin and Koopman (2012), for example, cannot be used and need to be modified. In particular, we provide three new results.

First, we apply the iterative conditional mode algorithm of Besag (1986) for obtaining the posterior modes of the loadings and the factors, simultaneously. The algorithm iterates between the updating of the loadings conditional on the factors and vice versa. We show that this algorithm can be implemented in a computationally efficient manner by exploiting the results of Jungbacker and Koopman (2015) and Mesters and Koopman (2014). We show that after convergence, we have obtained the joint posterior mode of the loadings and factors.

Second, we develop a two-step estimation procedure for the deterministic vector of model parameters using

likelihood-based methods. In the first step, we treat the elements of the loading matrix as deterministic and estimate these and remaining parameters in the model using standard state-space methods (see Doz et al., 2012, and Jungbacker & Koopman, 2015). This step produces maximum likelihood estimates for the loadings and parameters that pertain to the distributions of the factors and the disturbances. In the second step, we estimate the parameters that pertain to the distribution of the loadings in which the maximum likelihood estimates of the loadings from the first step are used as observations.

Third, we consider the estimation of other posterior statistics, such as the posterior mean and the posterior variance of the loadings and the factors. We argue that analytical solutions are not available when both $\lambda_i$ and $\alpha_t$ are considered stochastic (see also the arguments provided in Bishop (2006, chaps. 8, 12). We can resort to simulation methods, a standard solution for the estimation of latent variables in nonlinear models. However, given the typical large dimensions of the dynamic factor model ($N, T > 100$), standard simulation methods converge slowly and are unreliable because they are subject to so-called infinite variance problems (see Geweke, 1989). To solve this problem, we factorize the estimation into two separate parts: one for the loadings and one for the factors. We show that the integral over the factors can be calculated analytically, while the integral over the remaining $\lambda$-dependent terms can be calculated using basic simulation methods. The performance of this "integrated" simulation-based estimation method is more stable and has overall good properties.

The benefits of our model specification and estimation methods can be summarized as follows. First, our simulation study shows that the empirical Bayes joint posterior mode estimates for the common components $\lambda_i' \alpha_t$ are more accurate in the mean squared error (MSE) sense when compared to the maximum likelihood estimates. The differences are large and robust to changes in panel dimensions, the number of factors, and different sampling schemes for the loadings and error terms. The individual results for the loadings and factors are mixed and depend on the panel dimensions. For $N = T$, the loadings and the factors are estimated more accurately, with the empirical Bayes methods, but for $N \neq T$ only, either the loadings or the factors are estimated more accurately. Second, additional simulation results show that the relative gains in MSE for the common component increase when we include irrelevant and weak factors. These simulation settings are argued to be empirically relevant in, for example, Onatski (2012, 2015). Third, we show in our empirical study that the out-of-sample forecast errors of the empirical Bayes methods are smaller when compared to those resulting from the maximum likelihood estimates for a panel of macroeconomic and financial time series that was previously analyzed in Stock and Watson (2012). Fourth, from a computational perspective, by computing several integrals analytically, we reduce the computational complexity when compared to hierarchical Bayesian and Markov chain

Monte Carlo (MCMC) methods that also aim to learn about the prior distributions. While we predominantly focus on empirical Bayes methods, the results can be adapted for full Bayesian inference methods as well. Chan and Jeliazkov (2009) and McCausland, Miller, and Pelletier (2011) provide additional methods based on sparse matrix factorizations, which can increase the computational efficiency of MCMC methods. Finally, our methods produce posterior estimates that are not subject to errors and do not depend on Taylor expansions or discrete function approximations. Therefore, they compare favorably to other algorithms for nonlinear state-space models such as the extended Kalman filter of Anderson and Moore (1979) and the unscented Kalman filter of Julier and Uhlmann (1997).

The remainder of this paper is organized as follows. In the next section, we detail the specification of the dynamic factor model with stochastic loadings. In section III, we discuss the implementation of the empirical Bayesian estimation methods. The finite sample properties of the posterior mode estimates are studied in a Monte Carlo study that is presented in section IV. The methods are evaluated for different panel sizes and different numbers of factors. In section V, we present the results from our empirical study for macroeconomic forecasting with many predictors (see Stock & Watson, 2002b, 2012). Section VI concludes the paper and provides some directions for further research.

## II. Dynamic Factor Model with Stochastic Loadings

We define the dynamic factor model with stochastic loadings for $N$ variables indexed by $i = 1, \ldots, N$. The variables are observed over time for a span of $T$ periods; each time period is indexed by $t = 1, \ldots, T$. The observation vector $y_t = (y_{1,t}, \ldots, y_{N,t})'$ is modeled by

$$
\begin{aligned}
y_t &= \Lambda \alpha_t + \epsilon_t, & \epsilon_t &\sim NID(0, \Omega), \\
\Lambda &= (\lambda_1, \ldots, \lambda_N)', & \lambda_i &\sim NID(\delta, \Sigma_\lambda), \\
\alpha_{t+1} &= H\alpha_t + \eta_t, & \eta_t &\sim NID(0, \Sigma_\eta), \\
t &= 1, \ldots, T, & & (2)
\end{aligned}
$$

where $\Lambda$ is the $N \times r$ loading matrix, with $r < N$, $\alpha_t$ is the $r \times 1$ vector of common dynamic factors, and $\epsilon_t$ is the $N \times 1$ disturbance vector, with mean zero and variance matrix $\Omega$. The loading vectors $\lambda_i$ are normally and independently distributed with $r \times 1$ mean vector $\delta$ and $r \times r$ variance matrix $\Sigma_\lambda$. The dynamic factors $\alpha_t$ follow a vector autoregressive process of order 1, with $r \times r$ coefficient matrix $H$, and $\eta_t$ is an $r \times 1$ disturbance vector, with $r \times r$ variance matrix $\Sigma_\eta$. The initial state vector $\alpha_1$ is normally distributed with mean 0 and variance matrix $P_1$; more details of its treatment are given below. Our estimation methods are sufficiently general to consider more elaborate specifications for the factors. For notational and expositional convenience, we focus on the basic specification in equation (2).

The observation equation of model (2) can also be written as

$$y_t = (\alpha_t' \otimes I_N)\lambda + \epsilon_t, \qquad \epsilon_t \sim N(0, \Omega), \tag{3}$$

where $\lambda = \text{vec}(\Lambda) = (\acute{\lambda}_1', \ldots, \acute{\lambda}_r')'$ and the $N \times 1$ vector $\acute{\lambda}_j$ is the $j$th column of $\Lambda$, for $j = 1, \ldots, r$. This alternative representation of the observation equation is convenient for our exposition that follows.

To identify the factor space and the underlying parameters, several strategies can be pursued (see Bai & Li, 2012, and Chan, Leon-Gonzalez, & Strachan, forthcoming, for recent discussions). In this paper, we consider the following set of identifying restrictions.

**Assumption 1.** Consider the dynamic factor model in equation (2). We assume the following:

a. **Common factors.** The $r \times 1$ vector of common factors $\alpha_t$ is stationary and $\Sigma_\eta = I_r$. The initial state is given by $\alpha_1 \sim N(0, P_1)$, where $P_1 = (I_r - HH')^{-1}$ is the initial but also the unconditional variance of $\alpha_t$. The common innovations $\eta_t$ and the initial state $\alpha_1$ are mutually independent and distributed independently of the loading vectors $\lambda_i$ and the disturbances $\epsilon_{i,s}$, for all $i = 1, \ldots, N$ and $s, t = 1, \ldots, T$.

b. **Loading vectors.** The loading vectors $\lambda_i$ in equation (2) are distributed normally, with mean $\delta$ and positive definite variance $\Sigma_\lambda$, they are mutually independent, and they are distributed independently from the disturbances $\epsilon_t$, for all $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

c. **Disturbances.** The disturbance vectors $\epsilon_t$ in equation (2) are distributed normally and independently with mean 0 and diagonal variance matrix $\Omega = \text{diag}(\omega_1^2, \ldots, \omega_N^2)$.

Assumptions 1a and 1c are standard for exact dynamic factor models and identify the factor space, subject to an $r \times r$ orthonormal matrix (see, e.g., Doz et al., 2012, and Banbura & Modugno, 2014). Assumption 1b imposes the shrinkage restriction. Other shrinkage or sparse priors for the loadings are considered in, among others, Bhattacharya and Dunson (2011), Kaufmann and Schumacher (2013), and Nakajima and West (2013), but these prior specifications require full Bayesian estimation methods. Assumption 1 is sufficient for forecasting purposes. When we are interested in a particular rotation of the factors, we need to impose $r(r+1)/2$ additional restrictions on the loading matrix, which take the form of a lower triangular matrix with positive diagonal elements. In our simulation study, we treat this restricted $r \times r$ triangular matrix as deterministic and estimate the coefficients along with the other parameters. In our empirical application, we are interested in forecasting and do not impose these additional restrictions. Assumption 1c rules out cross-section and serial correlation in the disturbances but allows for heteroskedasticity in the cross-section dimension. Serial correlation can be handled by rewriting the model (see Stock & Watson, 2005, and Jungbacker and Koopman, 2015). We consider serially correlated disturbances in our simulation study in section IV. We emphasize that we view the dynamic factor model under assumption 1 as a restricted version of the approximate factor model considered in, for example, Stock and Watson (2002a) and Doz et al. (2012).

Equation (2) defines the dynamic factor model with stochastic loadings. The parameters pertaining to the model are collected in the vector $\psi$. For our model (2), it contains the unrestricted elements of $\delta$, $\Sigma_\lambda$, $H$, and $\Omega$. We notice that the number of deterministic parameters is much smaller compared to the standard dynamic factor model with deterministic loadings as considered in, for example, Jungbacker and Koopman (2015) and Doz et al. (2012). When $\Sigma_\lambda \to \infty$, the prior restriction on the loading matrix vanishes.

In the exposition that follows, we assume that the panel of observations is balanced and there are no missing observations. This assumption is merely for notational convenience; in the online appendix, we present the adjustments for the methods of section III for the case of missing observations in the panel. The adjustments are minor.

## III. Estimation of the Loadings, Factors, and Parameters

We develop the parametric empirical Bayesian methods for the estimation of the loadings, factors, and the parameter vector for the dynamic factor model in equation (2) given the conditions in assumption 1. The posterior modes of the loadings and factors are estimated using filtering methods, and the vector of deterministic parameters is estimated by maximum likelihood. We provide an iterative algorithm for computing the joint posterior modes of the loadings and factors. The posterior mode estimates provide the point estimates and are the subject of our Monte Carlo study in section IV. Further, we develop methods for estimating other posterior statistics for the loadings and the factors. These are necessary for the construction of finite sample confidence intervals.

### A. Joint Posterior Mode Analysis

The posterior modes for $\lambda = \text{vec}(\Lambda)$ and $\alpha = (\alpha_1', \ldots, \alpha_T', \alpha_{T+1}')'$ are defined by

$$\{\hat{\lambda}, \hat{\alpha}\} = \arg\max_{\lambda, \alpha} \log p(\lambda, \alpha | y; \psi), \tag{4}$$

where $y = (y_1', \ldots, y_T')'$. The direct optimization of $p(\lambda, \alpha | y; \psi)$ with respect to $\lambda$ and $\alpha$ is complicated as the first-order conditions for $\lambda$ and $\alpha$ depend on each other and solving analytically, or numerically, for either one is infeasible when $N$ and $T$ are large. The following theorem shows that we can separate the first-order conditions of the posterior density into two parts: one part that can be used for calculating the posterior mode of $\lambda$ and another part that can be used for calculating the posterior mode of $\alpha$.

**Theorem 1.** *For $y$ defined by model* (2), *it holds, under assumption 1, that*

$$\left.\frac{\partial \log p(\lambda, \alpha | y; \psi)}{\partial(\lambda', \alpha')'}\right|_{\lambda=\tilde{\lambda}, \alpha=\tilde{\alpha}}$$

$$= \left.\frac{\partial \log\left[p(\lambda | y; \alpha = \tilde{\alpha}, \psi) p(\alpha | y; \lambda = \tilde{\lambda}, \psi)\right]}{\partial(\lambda', \alpha')'}\right|_{\lambda=\tilde{\lambda}, \alpha=\tilde{\alpha}},$$

*for all given $\tilde{\lambda}$ and $\tilde{\alpha}$.*

The proof is given in the online appendix. The decomposition of the score function in theorem 1 enables the efficient computation of the posterior modes of both $\lambda$ and $\alpha$. It holds that $\hat{\lambda}$ can be found by maximizing $\log p(\lambda | y; \alpha = \hat{\alpha}, \psi)$ with respect to $\lambda$. Also, $\hat{\alpha}$ can be found by maximizing $\log p(\alpha | y; \lambda = \hat{\lambda}, \psi)$ with respect to $\alpha$. The modes, means, and variances of $p(\lambda | y; \alpha = \hat{\alpha}, \psi)$ and $p(\alpha | y; \lambda = \hat{\lambda}, \psi)$ can be evaluated using standard methods. In particular, conditionally on $\alpha$ the observation equation (3), together with the marginal density for $\lambda$, forms a linear gaussian regression model. Multivariate regression methods are used to evaluate the mean and variance of $p(\lambda | y; \alpha = \hat{\alpha}, \psi)$. The mode follows from the equality between the mean and mode for the gaussian density. Also, conditional on $\lambda$ model (2) is a linear gaussian state-space model. The mean, and thus the mode, and variance of $p(\alpha | y; \lambda = \hat{\lambda}, \psi)$ are evaluated by the Kalman filter smoother (see Durbin & Koopman, 2012).

However, we can obtain the modes $\hat{\lambda}$ or $\hat{\alpha}$ only when we have knowledge of either $\hat{\alpha}$ or $\hat{\lambda}$. The iterative conditional mode algorithm proposed in Besag (1986) provides a simple and stable solution for this problem. For model (2), the following theorem summarizes this algorithm.

**Theorem 2.** *Suppose that $p(\lambda, \alpha | y; \psi)$ is unimodal in $\lambda$ and $\alpha$, with $\hat{\lambda}$ and $\hat{\alpha}$ being the only stationary points. We assume that the conditions in assumption 1 hold. For given arbitrary starting values $\lambda^{(s)} \neq \mathbf{0}$, for $s = 0$, and using theorem 1, the posterior modes $\hat{\lambda}$ and $\hat{\alpha}$ in equation (4) can be obtained by iterating between*

*i.* $\alpha^{(s)} = \mathrm{E}(\alpha | y; \lambda = \lambda^{(s-1)}, \psi)$
*ii.* $\lambda^{(s)} = \mathrm{E}(\lambda | y; \alpha = \alpha^{(s)}, \psi)$
*iii.* $s = s + 1$

*until convergence.*

The proof is presented in the online appendix, together with the details for the efficient computation of $\mathrm{E}(\alpha | y; \lambda = \lambda^{(s-1)}, \psi)$ and $\mathrm{E}(\lambda | y; \alpha = \alpha^{(s)}, \psi)$. The computational advances are based on Jungbacker and Koopman (2015) and Mesters and Koopman (2014). The algorithm can be viewed as the expectation conditional maximization (ECM) algorithm of Meng and Rubin (1993), where the E-step is unity. We notice that the steps i and ii in this algorithm are M-steps since, given the gaussian assumption 1, the conditional

expectations are equal to conditional maximization steps. The assumption of unimodality is not different from the assumptions on the ECM algorithm and, as argued in Meng and Rubin (1993), not different from the assumptions for the EM algorithm. In practice, we use as a convergence criteria $||\lambda_{i,j}^{(s)} / \lambda_{i,j}^{(s-1)} - 1|| < 10^{-5}$ and $||\alpha_{j,t}^{(s)} / \alpha_{j,t}^{(s-1)} - 1|| < 10^{-5}$, for all $i = 1, \ldots, N, j = 1, \ldots, r$, and $t = 1, \ldots, T$.

While the conditional modes can be calculated iteratively up to any degree of accuracy, we should emphasize that the curvatures around $p(\lambda, \alpha | y; \psi)$ and $p(\lambda | y; \alpha = \hat{\alpha}, \psi) p(\alpha | y; \lambda = \hat{\lambda}, \psi)$ are different. The former takes into account the posterior dependence between $\lambda$ and $\alpha$, while the latter does not. Nevertheless, the point estimates for the factors in equation (4) are interesting from a classical perspective as they can be compared to the standard maximum likelihood estimates that compute the posterior mean of the factors given the maximum likelihood estimate for the loading matrix. To operationalize the computation of the posterior mode estimates using theorem 2, we require a value for the parameter vector $\psi$; its estimation is the topic of the next section.

### B. Likelihood Evaluation

The parameter vector $\psi$ contains parameters that are part of the distributions of the loadings, factors, and disturbances. The estimation of the parameter vector is based on maximum likelihood. For model (2), it is difficult to evaluate the marginal likelihood $L(\psi; y) = p(y; \psi)$ analytically. This follows as the product of stochastic variables $\Lambda$ and $\alpha_t$ prohibits closed-form solutions for the integral representation of the marginal likelihood given by

$$p(y; \psi) = \int_{\alpha} \int_{\lambda} p(y, \lambda, \alpha; \psi) \, d\lambda \, d\alpha. \tag{5}$$

More specifically, sequential methods (such as filtering) and iterative estimation methods (such as ECM) require at some point the evaluation of the conditional mean function $\mathrm{E}(\Lambda \alpha_t | y_1, \ldots, y_s; \psi)$, for some $s \in \{1, \ldots, T\}$. As closed-form expressions do not exist for the conditional expectation of products of stochastic variables, we cannot use these methods.

To solve this problem, we can rely on a two-step estimation method. We decompose the parameter vector $\psi = \{\psi_1, \psi_2\}$, where $\psi_1 = \{\Omega, H\}$ and $\psi_2 = \{\delta, \Sigma_{\lambda}\}$. In the first step, we consider the classical state-space problem given by

$$\{\tilde{\lambda}, \hat{\psi}_1\} = \arg\max_{\lambda, \psi_1} \log p(y | \lambda; \psi_1), \tag{6}$$

where $\lambda$ is treated as deterministic (see Jungbacker & Koopman, 2015, and Doz et al., 2012). The conditional likelihood $p(y | \lambda, \psi_1)$ is evaluated via the prediction error decomposition given by

$$\log p(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi}_1) = -\frac{NT}{2} \log 2\pi$$
$$-\frac{1}{2} \sum_{t=1}^{T} \left( \log |\boldsymbol{F}_t| + \boldsymbol{v}_t' \boldsymbol{F}_t^{-1} \boldsymbol{v}_t \right), \qquad (7)$$

where the quantities $\boldsymbol{v}_t = \boldsymbol{y}_t - \boldsymbol{\Lambda}\boldsymbol{a}_t$ and $\boldsymbol{F}_t = \boldsymbol{\Lambda}\boldsymbol{P}_t\boldsymbol{\Lambda}' + \boldsymbol{\Omega}$ are computed by the Kalman filter. It holds that $\boldsymbol{a}_t = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{y}_{t-1},\ldots,\boldsymbol{y}_1,\boldsymbol{\lambda})$ and $\boldsymbol{P}_t = \mathrm{Var}(\boldsymbol{\alpha}_t|\boldsymbol{y}_{t-1},\ldots,\boldsymbol{y}_1,\boldsymbol{\lambda})$, which are also functions of $\boldsymbol{\lambda}$ and are computed by the Kalman filter. The likelihood can be optimized with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}_1$ using numerical optimization methods. Computation efficiency can be improved by using the methods discussed in Jungbacker and Koopman (2015). We emphasize that $\tilde{\boldsymbol{\lambda}}$ is the maximum likelihood estimate for the loadings.

In the second step, we estimate the parameters $\boldsymbol{\psi}_2$ by solving

$$\hat{\boldsymbol{\psi}}_2 = \arg\max_{\boldsymbol{\psi}_2} \log p(\tilde{\boldsymbol{\lambda}}; \boldsymbol{\psi}_2), \qquad (8)$$

where the observations $\tilde{\boldsymbol{\lambda}}$ are obtained from the first step, and the density $p(\tilde{\boldsymbol{\lambda}}; \boldsymbol{\psi}_2)$ is implied by assumption 1b. Given that the first step yields consistent estimates as shown in Bai and Li (2012), it is easy to verify that the second step gives consistent estimates for the parameter vector $\boldsymbol{\psi}_2$ (see, e.g., Newey & McFadden, 1994). If certain elements of $\boldsymbol{\lambda}$ are fixed to 0 for identification purposes, these are not included for estimation in the second step.

### C. Posterior Statistics for the Loadings

We consider the evaluation of the posterior means and variances of the loadings and the factors. For the mean and variance, analytical solutions are not available as the product of $\boldsymbol{\Lambda}$, and $\boldsymbol{\alpha}_t$ prohibits closed-form solutions. We show that when evaluating the mean and variance of the loadings and factors, the factors can be integrated out analytically. To evaluate the resulting expressions, which depend only on the latent vector $\boldsymbol{\lambda}$, we develop adequate importance densities based on the Laplace approximation (see So, 2003, and Jungbacker & Koopman, 2007). The importance densities are used for integration via simulation. Hence, the posterior evaluation is done partly analytically and partly via simulation.

Let $\boldsymbol{f}(\boldsymbol{\lambda})$ denote some arbitrary vector function of $\boldsymbol{\lambda}$. In the simplest case, $\boldsymbol{f}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}$. We are interested in estimating the conditional mean function $\bar{\boldsymbol{f}} = \mathrm{E}(\boldsymbol{f}(\boldsymbol{\lambda})|\boldsymbol{y}; \boldsymbol{\psi})$. It holds that

$$\bar{\boldsymbol{f}} = \int_{\boldsymbol{\lambda}} \boldsymbol{f}(\boldsymbol{\lambda}) p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}, \qquad (9)$$

where the posterior density $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$ is complicated and is not known analytically. We notice that $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \propto p(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi}) p(\boldsymbol{\lambda}; \boldsymbol{\psi})$, where $p(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi})$ is provided by the

Kalman filter via the prediction error decomposition, equation (7). The elements of the latent loading vector $\boldsymbol{\lambda}$ enter the log density, equation (7), nonlinearly in both the mean vector and the variance matrix. Therefore, the posterior density $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$ has a complicated form.

For the numerical evaluation of the integral in equation (9), we make use of the importance sampling technique. We can express the integral, equation (9), in terms of the importance density $g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$ and obtain

$$\bar{\boldsymbol{f}} = \int_{\boldsymbol{\lambda}} \boldsymbol{f}(\boldsymbol{\lambda}) \frac{p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})}{g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})} g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}. \qquad (10)$$

In the online appendix we show that the Monte Carlo estimate of equation (10) is given by

$$\hat{\bar{\boldsymbol{f}}} = \frac{\sum_{j=1}^{M} \boldsymbol{f}(\boldsymbol{\lambda}^{(j)}) w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}^{(j)}; \boldsymbol{\psi})}{\sum_{j=1}^{M} w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}^{(j)}; \boldsymbol{\psi})}, \qquad (11)$$

for a sufficiently large integer $M$, where $w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}^{(j)}; \boldsymbol{\psi}) = p(\boldsymbol{y}|\boldsymbol{\lambda}^{(j)}; \boldsymbol{\psi}) / g(\boldsymbol{y}|\boldsymbol{\lambda}^{(j)}; \boldsymbol{\psi})$ and where the samples $\boldsymbol{\lambda}^{(j)}$ are drawn from $g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$. The importance density $g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$ targets $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$, which is the marginal posterior density of the loadings that does not depend on $\boldsymbol{\alpha}$. An adequate importance density $g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$ that accurately approximates the marginal posterior density $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$ needs to be obtained. An initial candidate that may seem appropriate is $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}, \boldsymbol{\psi})$. While the location of this density is accurate, it turns out that its variance is too small relative to $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$ since $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}, \boldsymbol{\psi})$ does not account for the variance of the posterior mode estimate $\hat{\boldsymbol{\alpha}}$.

Instead, we choose $g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$ to follow a gaussian distribution, where the mean and the variance are equal to the mode and the curvature around the mode of $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$. So (2003) and Jungbacker and Koopman (2007) argue that the mode can be obtained by maximizing $\log p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) = \log p(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi}) + \log p(\boldsymbol{\lambda}; \boldsymbol{\psi}) - \log p(\boldsymbol{y}; \boldsymbol{\psi})$ with respect to $\boldsymbol{\lambda}$. We notice that the resulting conditional mode is different from the conditional mode given in theorem 2. From Jungbacker and Koopman (2007, theorem 1), it follows that the mode and the curvature around the mode of $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$ can be found by the following Newton-Raphson steps:

i. Initialize $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$.
ii. Compute

$$\boldsymbol{x} = \boldsymbol{\lambda}^* - \left\{ \ddot{p}(\boldsymbol{y}|\boldsymbol{\lambda}^*; \boldsymbol{\psi}) \right\}^{-1} \dot{p}(\boldsymbol{y}|\boldsymbol{\lambda}^*; \boldsymbol{\psi}),$$
$$\boldsymbol{C} = - \left\{ \ddot{p}(\boldsymbol{y}|\boldsymbol{\lambda}^*; \boldsymbol{\psi}) \right\}^{-1},$$

where

$$\dot{p}(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi}) = \frac{\partial \log p(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi})}{\partial \boldsymbol{\lambda}},$$
$$\ddot{p}(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi}) = \frac{\partial^2 \log p(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'}.$$

iii. Update $\boldsymbol{\lambda}^*$ by computing $\mathrm{E}(\boldsymbol{\lambda}|\boldsymbol{x};\boldsymbol{\psi})$, where the "observation" vector $\boldsymbol{x}$ is assumed to come from the model $\boldsymbol{x} = \boldsymbol{\lambda} + \boldsymbol{u}$, with $\boldsymbol{u} \sim N(0,\boldsymbol{C})$.

iv. Iterate between steps ii and iii until convergence.

After convergence, $\boldsymbol{\lambda}^*$ is equal to the mode of $p(\boldsymbol{\lambda}|y;\boldsymbol{\psi})$, and $\boldsymbol{C}$ is equal to the curvature around the mode. The main difficulty in the implementation of these steps is the computation of the derivatives in step ii since $\boldsymbol{\lambda}$ occurs in the log determinant, the inverse, and the predictive mean and variance of the density $p(y|\boldsymbol{\lambda};\boldsymbol{\psi})$ in equation (7). Given that the derivatives of the marginal likelihood are equivalent to the derivatives of the expected complete likelihood, Koopman and Shephard (1992) show that the derivatives can be computed in closed form. Explicit expressions for the dynamic factor model with given loadings are given in Banbura and Modugno (2014) and Jungbacker and Koopman (2015).

The resulting importance density is chosen as $g(\boldsymbol{\lambda}|y;\boldsymbol{\psi}) \equiv N(\boldsymbol{\lambda}^*,\boldsymbol{C}^*)$, where $\boldsymbol{C}^* = \mathrm{Var}(\boldsymbol{\lambda}) - \mathrm{Var}(\boldsymbol{\lambda})(\mathrm{Var}(\boldsymbol{\lambda}) + \boldsymbol{C})^{-1}\mathrm{Var}(\boldsymbol{\lambda})$ evaluated at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$. The observation model for $\boldsymbol{x}$ in step iii can be used for sampling and to compute the weights $w_\lambda(y,\boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})$. We typically ignore the off-diagonal elements in $\boldsymbol{C}$ as they do not affect the quality of the weights. In the online appendix, we show that the weights that result from this choice for the importance density have finite variance for all model specifications under consideration. Since the model for $\boldsymbol{x}$ is simple and static by construction, sampling from $g(\boldsymbol{\lambda}|y;\boldsymbol{\psi})$ can be implemented in a computationally efficient way.

By integrating out $\boldsymbol{\alpha}$ analytically via Kalman filter methods, we reduce our reliance on simulation and therefore the simulation variance. We can regard this as a direct application of the Rao-Blackwellization principle, which is discussed in general in Doucet, de Freitas, and Gordon (2001) and Durbin and Koopman (2012).

### D. Posterior Statistics for the Factors

Similar results can be obtained for functions of the factors. In particular, define $\boldsymbol{h}(\boldsymbol{\alpha})$ as a vector function of $\boldsymbol{\alpha}$. We are interested in estimating the conditional mean function $\bar{\boldsymbol{h}} = \mathrm{E}(\boldsymbol{h}(\boldsymbol{\alpha})|y;\boldsymbol{\psi})$. Similar to the loadings, it holds that

$$\bar{\boldsymbol{h}} = \int_\alpha \boldsymbol{h}(\boldsymbol{\alpha})p(\boldsymbol{\alpha}|y;\boldsymbol{\psi})\,\mathrm{d}\boldsymbol{\alpha}, \tag{12}$$

where the marginal conditional density $p(\boldsymbol{\alpha}|y;\boldsymbol{\psi})$ is unknown when $\boldsymbol{\lambda}$ is stochastic. We do not want to construct an additional importance density for the estimation of the factors for the following two reasons. First, when $\boldsymbol{\alpha}$ constitutes a persistent dynamic process, the accurate approximation of $p(\boldsymbol{\alpha}|y;\boldsymbol{\psi})$ in high dimensions is complicated (see, e.g., Koopman, Lucas, & Scharth, 2015). Second, even when an accurate importance density is obtained, sampling from it is typically computationally demanding. Therefore, we take a different approach for the evaluation of $\boldsymbol{h}(\boldsymbol{\alpha})$ and rewrite the integral in equation (12) as

$$\bar{\boldsymbol{h}} = \int_\lambda \int_\alpha \boldsymbol{h}(\boldsymbol{\alpha})p(\boldsymbol{\alpha},\boldsymbol{\lambda}|y;\boldsymbol{\psi})\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{\lambda}$$

$$= \int_\lambda \int_\alpha \boldsymbol{h}(\boldsymbol{\alpha})p(\boldsymbol{\alpha}|\boldsymbol{\lambda},y;\boldsymbol{\psi})p(\boldsymbol{\lambda}|y;\boldsymbol{\psi})\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{\lambda}$$

$$= \int_\lambda \mathrm{E}(\boldsymbol{h}(\boldsymbol{\alpha})|y,\boldsymbol{\lambda};\boldsymbol{\psi})\,p(\boldsymbol{\lambda}|y;\boldsymbol{\psi})\,\mathrm{d}\boldsymbol{\lambda}. \tag{13}$$

We have rewritten the conditional mean function in terms of the marginal posterior density of the loadings. For a given value of $\boldsymbol{\lambda}$, we can evaluate $\mathrm{E}(\boldsymbol{h}(\boldsymbol{\alpha})|y,\boldsymbol{\lambda};\boldsymbol{\psi})$ using the Kalman filter smoother for many functions $\boldsymbol{h}(\boldsymbol{\alpha})$. From section IIIB, we have learned that analytic expressions for the moments of $p(\boldsymbol{\lambda}|y;\boldsymbol{\psi})$ do not exist. In the online appendix, we derive the Monte Carlo estimate for $\bar{\boldsymbol{h}}$, which is given by

$$\hat{\bar{\boldsymbol{h}}} = \frac{M^{-1}\sum_{j=1}^{M}\mathrm{E}(\boldsymbol{h}(\boldsymbol{\alpha})|y,\boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})\,w_\lambda(y,\boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})}{M^{-1}\sum_{j=1}^{M}w_\lambda(y,\boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})}, \tag{14}$$

for some large integer $M$, where the weights $w_\lambda(y,\boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})$ are defined below equation (11) and the samples $\boldsymbol{\lambda}^{(j)}$ are drawn from $g(\boldsymbol{\lambda}|y;\boldsymbol{\psi})$. Given that for every draw $\boldsymbol{\lambda}^{(j)}$ we can evaluate $\mathrm{E}(\boldsymbol{h}(\boldsymbol{\alpha})|y,\boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})$ analytically, we require only the construction of the importance density $g(\boldsymbol{\lambda}|y;\boldsymbol{\psi})$ (see the discussion in section IIIC).

## IV. Simulation Study

In this section we study the finite sample properties of the methods discussed in section III. Our main interest is to compare the empirical Bayes estimates with the classical maximum likelihood estimates. The empirical Bayes point estimates that we consider are the posterior mode estimates from section IIIA. The estimates are computed given the estimated parameter vector $\hat{\boldsymbol{\psi}}$, which is obtained as discussed in section IIIB. The maximum likelihood estimates for the loadings are given by $\tilde{\boldsymbol{\lambda}}$ in equation (6), and the corresponding smoothed estimates for the factors are given by $\tilde{\boldsymbol{\alpha}}_t = \mathrm{E}(\boldsymbol{\alpha}_t|y;\tilde{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}_1)$ (see Doz et al., 2012, and Jungbacker & Koopman, 2015). We study the differences between the empirical Bayes and maximum likelihood estimates for different data-generating processes that are outlined below.

### A. Simulation Design

We study the dynamic factor model for different cross-section and time series dimensions. We include combinations for $N,T = 50,100$, as well as the relevant combination $N = 150$ and $T = 200$, which is typically encountered in macroeconomic panels. The number of factors is chosen to be equal to $r = 3$ or $r = 5$. During the simulation study, we assume that the true number of factors is known unless otherwise indicated. In empirical applications, it is possible to rely on economic theory, information criteria, or hypothesis tests to determine an appropriate number of factors (see, e.g., Bai & Ng, 2002, and Onatski, 2009).

We draw $S = 500$ different panels of observations from model (2) for each combination of panel size and number of factors. We denote the sampled vectorized panels by $\boldsymbol{y}(s)$ for $s = 1, \ldots, S$. Each observation vector has its own "true" loadings and factors, $\boldsymbol{\lambda}(s)$ and $\boldsymbol{\alpha}(s)$. The loading vectors $\boldsymbol{\lambda}_i(s)$ are drawn from a variety of mixture distributions. In particular, we draw the elements of the loading vector from

$$\pi(\lambda_{i,j}) = k_1 N(\mu_1, \sigma_1^2) + \ldots + k_s N(\mu_s, \sigma_s^2), \tag{15}$$

where the values for $\mu_n$, $\sigma_n^2$, $k_n$, for $n = 1, \ldots, s$ and $s$ are taken such that the loadings have normal, trimodal, skewed, and kurtotic distributions. The values for the loading settings are given in Marron and Wand (1992, table 1). The variety of sampling schemes for the "true" loadings ensures that our results do not depend on the normality assumption for the loadings. Additional to assumption 1, we restrict the loading matrix $\boldsymbol{\Lambda} = [\boldsymbol{\Lambda}_1' : \boldsymbol{\Lambda}_2']'$, such that the $r \times r$ matrix $\boldsymbol{\Lambda}_1$ is lower triangular with positive diagonal elements. This identifies a particular rotation for the factors and allows us to calculate mean squared error statistics for the loading and factor estimates. We estimate the unrestricted elements of the matrix $\boldsymbol{\Lambda}_1$ together with the other deterministic parameters of the model and keep them fixed in the posterior mode algorithm of section IIIA.

The remaining details of the simulation design are similar to those in Doz et al. (2012). In particular, the dynamic factors are simulated from the autoregressive process in equation (2) with autoregressive polynomial matrix $\boldsymbol{H}$ and variance matrix $\boldsymbol{\Sigma}_\eta = \boldsymbol{I}_r$, such that $\text{Var}(\boldsymbol{\alpha}_t) = (\boldsymbol{I}_r - \boldsymbol{H}\boldsymbol{H}')^{-1}$. The elements for the diagonal of $\boldsymbol{H}$ are drawn from $U(0.5, 0.9)$. The off-diagonal elements are set to 0. For both the maximum likelihood and the empirical Bayes methods, the factor process is correctly specified.

The error term $\boldsymbol{\epsilon}_t$ incorporates cross-sectional and serial correlation. Given draws $\lambda_{i,j}$, $\boldsymbol{\epsilon}_t$ is generated by

$$\begin{aligned} \boldsymbol{\epsilon}_t &= \gamma \boldsymbol{\epsilon}_{t-1} + \boldsymbol{\zeta}_t, & \boldsymbol{\zeta}_t &\sim N(\boldsymbol{0}, \boldsymbol{\mathcal{T}}), \\ \alpha_i &= \frac{\beta_i}{1 - \beta_i} \boldsymbol{\lambda}_i'(\boldsymbol{I}_r - \boldsymbol{H}\boldsymbol{H}')^{-1} \boldsymbol{\lambda}_i, & \beta_i &\sim U(u, 1 - u), \\ \mathcal{T}_{ij} &= \tau^{|i-j|} (1 - \gamma^2) \sqrt{\alpha_i \alpha_j}, & i, j &= 1, \ldots, N, \end{aligned} \tag{16}$$

where $\gamma$ governs the serial correlation in the disturbances. We take $\gamma = 0.5$. The coefficient $\beta_i$ captures the ratio between the variance of $\epsilon_{i,t}$ and the total variance of $y_{i,t}$. We draw $\beta_i$ from the uniform distribution $U(u, 1 - u)$ with $u = 0.1$, which sets the ratio between the variances to 0.5 on average. Finally, the magnitude of the cross-sectional correlation is governed by $\tau$, which we set equal to 0.5. We emphasize that we have implemented our methods of section III under the conditions of assumption 1. By result, the error process is misspecified as one would expect in a typical macroeconomic application.

## B. Comparisons of Empirical Bayes versus Maximum Likelihood

We study the relative accuracy of the empirical Bayes estimates to the maximum likelihood estimates. For the comparison, we consider the mean squared errors given by

$$\text{MSE}(\check{\boldsymbol{\lambda}}) = S^{-1} \sum_{s=1}^{S} (Nr)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{r} \left( \lambda_{i,j}(s) - \check{\lambda}_{i,j}(s) \right)^2,$$

$$\text{MSE}(\check{\boldsymbol{\alpha}}) = S^{-1} \sum_{s=1}^{S} (Tr)^{-1} \sum_{t=1}^{T} \sum_{j=1}^{r} \left( \alpha_{j,t}(s) - \check{\alpha}_{j,t}(s) \right)^2,$$

$$\text{MSE}(\check{\boldsymbol{\lambda}}' \check{\boldsymbol{\alpha}}) = S^{-1} \sum_{s=1}^{S} (NT)^{-1}$$
$$\times \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \boldsymbol{\lambda}_i'(s) \boldsymbol{\alpha}_t(s) - \check{\boldsymbol{\lambda}}_i'(s) \check{\boldsymbol{\alpha}}_t(s) \right)^2, \tag{17}$$

where $\check{\boldsymbol{\lambda}} \in \{\hat{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\lambda}}\}$ and $\check{\boldsymbol{\alpha}} \in \{\hat{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\alpha}}\}$; these definitions also apply to their elements. For different data-generating processes that are discussed in section IVA, we present the relative mean squared errors (rMSE) between the empirical Bayes and maximum likelihood mean squared errors. In particular, we present $\text{rMSE}(\lambda) = \text{MSE}(\hat{\boldsymbol{\lambda}})/\text{MSE}(\tilde{\boldsymbol{\lambda}})$, $\text{rMSE}(\alpha) = \text{MSE}(\hat{\boldsymbol{\alpha}})/\text{MSE}(\tilde{\boldsymbol{\alpha}})$, and $\text{rMSE}(\lambda'\alpha) = \text{MSE}(\hat{\boldsymbol{\lambda}}'\hat{\boldsymbol{\alpha}})/\text{MSE}(\tilde{\boldsymbol{\lambda}}'\tilde{\boldsymbol{\alpha}})$, which are smaller than 1 whenever the MSE of the empirical Bayes methods is smaller.

The results of our simulation study for the different data-generating processes described in section IVA are summarized in panels A to C of table 1. Based on the relative MSE statistics, we make four observations. First, the relative MSE of the loadings, panel A, shows that for $N = T$ and $N > T$, the empirical Bayes methods are more accurate. When $T > N$, the maximum likelihood method gives more accurate estimates. The empirical Bayes methods improve relative to the maximum likelihood estimates when the number of factors is increased from $r = 3$ to $r = 5$. The relative differences between the two methods show gains of up to 25% and losses up to 10% for $T > N$ and $r = 3$. Second, for the factors, panel B, we find that the empirical Bayes methods are more accurate for $N = T$ and $T > N$, whereas for $N > T$, the maximum likelihood method is more accurate. Overall, the relative differences between the empirical Bayes and maximum likelihood methods for the factors are smaller when compared to those for the loadings. This is not surprising since the factors have the same specification. We conclude that for $N = T$, the empirical Bayes estimates perform well for both the factors and the loadings. For $N \neq T$, the results are mixed; either the loadings or the factors are estimated with lower MSE by the empirical Bayes methods. Third, since there is a trade-off between the accuracy of the loadings and the factors for $N \neq T$, we show the MSE statistics for the common components in panel C. For this case, the relative MSE is always in favor of the empirical Bayes estimates. The gains for the common components are between

TABLE 1.—SIMULATION RESULTS FOR THE EMPIRICAL BAYES AND MAXIMUM LIKELIHOOD ESTIMATES

| | | | A. rMSE($\lambda$) | | | | | B. rMSE($\alpha$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | T | r | L = 1 | L = 2 | L = 3 | L = 4 | L = 5 | L = 1 | L = 2 | L = 3 | L = 4 | L = 5 |
| 50 | 50 | 3 | 0.885 | 0.911 | 0.896 | 0.863 | 0.886 | 0.975 | 0.976 | 0.986 | 0.979 | 0.963 |
| 50 | 100 | 3 | 1.123 | 1.094 | 1.125 | 1.073 | 1.119 | 0.981 | 0.979 | 0.967 | 0.953 | 0.980 |
| 100 | 50 | 3 | 0.836 | 0.858 | 0.840 | 0.783 | 0.843 | 1.295 | 1.293 | 1.327 | 1.233 | 1.269 |
| 100 | 100 | 3 | 0.929 | 0.939 | 0.934 | 0.882 | 0.929 | 0.987 | 0.985 | 0.998 | 0.983 | 0.976 |
| 150 | 200 | 3 | 1.038 | 1.033 | 1.042 | 1.008 | 1.041 | 0.999 | 0.993 | 0.995 | 0.981 | 1.000 |
| 50 | 50 | 5 | 0.866 | 0.902 | 0.872 | 0.871 | 0.860 | 0.979 | 0.979 | 0.986 | 0.989 | 0.974 |
| 50 | 100 | 5 | 1.082 | 1.063 | 1.076 | 1.047 | 1.049 | 0.915 | 0.912 | 0.912 | 0.903 | 0.908 |
| 100 | 50 | 5 | 0.772 | 0.804 | 0.777 | 0.745 | 0.764 | 1.229 | 1.227 | 1.253 | 1.204 | 1.214 |
| 100 | 100 | 5 | 0.897 | 0.911 | 0.902 | 0.865 | 0.889 | 0.984 | 0.983 | 0.989 | 0.984 | 0.976 |
| 150 | 200 | 5 | 1.002 | 1.002 | 1.006 | 0.962 | 1.002 | 0.956 | 0.955 | 0.954 | 0.948 | 0.958 |

| | | | C. rMSE($\lambda'\alpha$) | | | | | D. rMSE($\lambda'\alpha$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | L = 1 | L = 2 | L = 3 | L = 4 | L = 5 | L = 1 | L = 2 | L = 3 | L = 4 | L = 5 |
| 50 | 50 | 3 | 0.837 | 0.884 | 0.855 | 0.778 | 0.841 | 0.720 | 0.786 | 0.761 | 0.727 | 0.686 |
| 50 | 100 | 3 | 0.924 | 0.955 | 0.914 | 0.856 | 0.941 | 0.777 | 0.820 | 0.808 | 0.786 | 0.734 |
| 100 | 50 | 3 | 0.821 | 0.909 | 0.839 | 0.754 | 0.819 | 0.689 | 0.792 | 0.730 | 0.687 | 0.655 |
| 100 | 100 | 3 | 0.899 | 0.917 | 0.903 | 0.837 | 0.899 | 0.714 | 0.756 | 0.753 | 0.730 | 0.668 |
| 150 | 200 | 3 | 0.949 | 0.958 | 0.949 | 0.909 | 0.962 | 0.724 | 0.743 | 0.760 | 0.732 | 0.666 |
| 50 | 50 | 5 | 0.793 | 0.857 | 0.800 | 0.758 | 0.784 | 0.723 | 0.807 | 0.748 | 0.737 | 0.695 |
| 50 | 100 | 5 | 0.862 | 0.894 | 0.869 | 0.815 | 0.848 | 0.770 | 0.800 | 0.800 | 0.798 | 0.733 |
| 100 | 50 | 5 | 0.773 | 0.856 | 0.786 | 0.713 | 0.758 | 0.673 | 0.772 | 0.698 | 0.679 | 0.634 |
| 100 | 100 | 5 | 0.849 | 0.876 | 0.855 | 0.780 | 0.839 | 0.684 | 0.734 | 0.716 | 0.703 | 0.643 |
| 150 | 200 | 5 | 0.919 | 0.927 | 0.919 | 0.866 | 0.926 | 0.684 | 0.706 | 0.710 | 0.699 | 0.642 |

| | | | E. rMSE($\lambda'\alpha$) | | | | | F. rMSE($\lambda'\alpha$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | a = 0.1 | a = 1 | a = 2 | a = 3 | a = 5 | b = 0.1$^2$ | b = 0.25$^2$ | b = 0.5$^2$ | b = 0.75$^2$ | b = 1$^2$ |
| 50 | 50 | 3 | 0.982 | 0.939 | 0.895 | 0.883 | 0.861 | 0.751 | 0.787 | 0.817 | 0.829 | 0.837 |
| 50 | 100 | 3 | 1.005 | 0.956 | 0.949 | 0.941 | 0.925 | 0.829 | 0.885 | 0.915 | 0.922 | 0.924 |
| 100 | 50 | 3 | 0.969 | 0.909 | 0.865 | 0.855 | 0.832 | 0.731 | 0.778 | 0.808 | 0.818 | 0.821 |
| 100 | 100 | 3 | 0.988 | 0.956 | 0.939 | 0.920 | 0.890 | 0.802 | 0.855 | 0.886 | 0.890 | 0.899 |
| 150 | 200 | 3 | 0.986 | 0.982 | 0.967 | 0.960 | 0.943 | 0.863 | 0.914 | 0.939 | 0.948 | 0.949 |
| 50 | 50 | 5 | 0.982 | 0.883 | 0.871 | 0.870 | 0.864 | 0.752 | 0.772 | 0.788 | 0.792 | 0.793 |
| 50 | 100 | 5 | 0.982 | 0.941 | 0.942 | 0.940 | 0.929 | 0.790 | 0.818 | 0.846 | 0.859 | 0.862 |
| 100 | 50 | 5 | 0.947 | 0.858 | 0.844 | 0.836 | 0.831 | 0.713 | 0.737 | 0.760 | 0.770 | 0.772 |
| 100 | 100 | 5 | 0.976 | 0.915 | 0.895 | 0.887 | 0.874 | 0.754 | 0.805 | 0.838 | 0.848 | 0.849 |
| 150 | 200 | 5 | 0.989 | 0.961 | 0.981 | 0.956 | 0.950 | 0.814 | 0.877 | 0.908 | 0.917 | 0.919 |

The code L indicates 1; normal(0,1), 2; normal(0,0.04), 3; trimodal, 4; skewed, or 5; outlier distribution for the true loadings, $a$ indicates the shape of the gamma distribution that governs the signal-to-noise ratio and $b$ gives the variance of the weak factors. Panels A–C show the baseline comparisons based on the simulation design of section IVA, Panel D considers the same design but includes one or two irrelevant factors. Panel E shows results for different signal-to-noise ratios. Panel F varies the variance of the last or last two loading columns.

5% and 25% for all model specifications. The largest gains are found for cases where $N > T$, but also for $N < T$, we find consistent gains between 5% and 15%. Fourth, and finally, the different sampling schemes for the loadings do not seem to have large effects on the outcomes. Hence, the normal approximation imposed by the model specification for the loadings is able to capture different sampling schemes for the loadings such as trimodal, skewed, and kurtotic equally well.

Overall, we may conclude that the empirical Bayes estimates are more accurate when compared to the maximum likelihood estimates for the common components for many data-generating processes. The magnitude of the gains depends on the panel dimensions. For all computations in this study, we have written the code in the Ox programming language, version 7.00 of Doornik (2007). The implementation codes are made available for the programming environments of Ox and Matlab.

### C. Comparisons with Irrelevant and Weak Factors

Next, we compare the empirical Bayes and maximum likelihood estimates for the case where there are irrelevant or weak factors in the model. The relevance of these settings is discussed in Onatski (2012, 2015). We perform three Monte Carlo experiments that aim to investigate specific empirically relevant cases. In the first scenario, we generate the data panel with two factors and then estimate the factors and loadings using the model with $r = 3$ and $r = 5$ factors. In this setup, there are either one or three irrelevant factors. All other features of the simulation design remain as in section IVA. We compare the estimated common components with the true common components using equation (17) and again report the ratio between the empirical Bayes and the maximum likelihood statistics. This setting is similar to that in Onatski (2015), who studies the least squares estimator of the common component. The results for the common components are presented in panel D of table 1.

We find that the empirical Bayes estimates increase in accuracy relative to the maximum likelihood estimates. This holds for all combinations of panel dimensions and sampling schemes for the loadings. When comparing panels C and D, we find that the rMSE of the common components reduces between 5% and 20% across the different model

specifications. The gain remains large when the panel dimensions increase. For example, for $N = 150$ and $T = 200$, the rMSE drops from 0.949 to 0.724 when we consider the model with $r = 3$ factors.

In the second scenario, we modify the fraction of the variance captured by the common component relative to the total variance. In section IVA, we follow Doz et al. (2012) and control the signal-to-noise ratio using $d_i = \beta_i/(1 - \beta_i)$ (see equation [16]), with $\beta_i \sim U(0.1, 0.9)$. This specification can be approximated by having a gamma distribution with shape parameter $a = 2$ and scale parameter 1 for $d_i$. We consider this approximation and vary the shape parameter for $a = 0.1, 1, 2, 3, 5$. This corresponds to varying $\text{Var}(\lambda_i'\alpha_t)/\text{Var}(y_{i,t})$ by 0.9, 0.7, 0.5, 0.2, 0.1, respectively. We evaluate the influence of the signal-to-noise ratio for the case where the loadings are drawn from the standard normal distribution. Other sampling schemes for the loadings show the same results.

The results for the common components' MSE comparison for different $N, T$ combinations are presented in panel E of table 1. We find that the relative gain of the empirical Bayes estimates increases when the signal-to-noise ratio becomes lower, that is, when $a$ increases. When the common component captures on average 90% of the variance, the empirical Bayes estimates provide only modest relative gains. For $N = 50$ and $T = 100$, there is even a small loss for the empirical Bayes estimates. When the signal-to-noise ratio becomes smaller, the relative gain of the empirical Bayes estimates increases.

In the third and last scenario, we modify the variance of some of the loading columns. This has the effect that different factors explain different portions of the total variance. In particular, when the true number of factors is $r = 3$, we take the variance of the third loading column as $b = 0.1^2, 0.25^2, 0.5^2, 0.75^2$, or $1^2$. The loadings in the other columns are drawn from the standard normal distribution and all other features of the simulation design are as in section IVA. For $r = 5$, we modify the variances of the fourth and fifth columns using the same values for $b$. The results for the common components' MSE comparison for different $N, T$ combinations are presented in panel F of table 1. Notice that for $b = 1^2$, the results correspond to those in the first column of panel C.

For this experiment we also find that the empirical Bayes estimates outperform the maximum likelihood estimates even when the strength of selected factors decreases from equally strong to weak. For example, the difference in relative accuracy of the common components' estimates can vary up to 10% for a model with $r = 3$, when the variance of one factor changes from $b = 0.1^2$ to $b = 1$. It shows to some extent that the empirical Bayes methods are more able to empirically identify weak factors. Based on these results from panels D to F, we carefully conjecture that empirical Bayes methods can perform well in applications where the factors are weak or when irrelevant factors are included.

TABLE 2.—SUMMARY OF THE TIME SERIES INCLUDED IN THE EMPIRICAL APPLICATION

|   | Category | Number of Series (144) |
|---|---|---|
| A | GDP components | 16 |
| B | Industrial production | 14 |
| C | Employment | 20 |
| D | Unemployment rate | 7 |
| E | Housing starts | 6 |
| F | Inventories | 6 |
| G | Prices | 37 |
| H | Wages | 6 |
| I | Interest rates | 13 |
| J | Money | 8 |
| K | Exchange rates | 5 |
| L | Stock prices | 5 |
| M | Consumer expectations | 1 |

### D. Additional Simulation Results

We briefly discuss several findings that are presented in detail in the online appendix. First, we have compared the empirical Bayes and maximum likelihood estimates to the principal components estimates. Comparisons were based on the common components and showed that both the empirical Bayes and maximum likelihood estimates outperform the principal components estimates. This was already documented for maximum likelihood estimation by Doz et al. (2012) and Bai and Li (2012) using trace statistics. We confirm their conclusions using rMSE statistics.

Second, we complement the results in table 1 for observation disturbances with only serial correlation, only cross-sectional correlation, or neither serial correlation nor cross-sectional correlation. These results are very similar to those presented in panels A to C of table 1. This should not be surprising since the empirical Bayes and the maximum likelihood estimates are affected to the same extent by misspecification. Hence, the relative statistics should remain the same overall when the signal-to-noise ratio is kept constant.

### V. Macroeconomic Forecasting

Our empirical study is concerned with the comparison of the empirical Bayes and maximum likelihood approaches for dynamic factor models based on a quarterly U.S. macroeconomic data set. The key question is whether and to what extent the empirical Bayes methods improve out-of-sample forecasts when compared to the maximum likelihood methods. We consider the data set of Stock and Watson (2012), which includes $N = 144$ macroeconomic and financial time series. These series capture a large part of the available disaggregated macroeconomic and financial time series. Table 2 summarizes the categories for which the time series are included. There are thirteen different blocks, including large blocks for employment series and prices. Each block is indexed by a letter. From this data set, we construct stationary quarterly time series following the guidelines in the online appendix of Stock and Watson (2012). The resulting panel ranges from 1959-1 to 2008-4, with $T = 200$. The panel is

TABLE 3.—PARAMETER ESTIMATES FOR THE VECTOR AUTOREGRESSIVE COEFFICIENTS AND THE MEAN AND VARIANCE OF THE LOADINGS FOR THE FACTOR MODELS WITH $r = 5$ AND $r = 7$ FACTORS

| | Autoregressive Coefficients $\boldsymbol{H}$ | | | | | | | Eigenvalues | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | real | img |
| **Five factors** | | | | | | | | | |
| | 0.911 | 0.065 | −0.161 | −0.287 | −0.094 | | | 0.945 | 0.000 |
| | 0.049 | 0.656 | 0.364 | 0.103 | −0.183 | | | 0.811 | 0.110 |
| | 0.014 | 0.278 | 0.555 | 0.090 | 0.048 | | | 0.811 | −0.110 |
| | 0.064 | −0.331 | −0.056 | 0.454 | −0.532 | | | 0.445 | 0.000 |
| | 0.016 | 0.009 | −0.059 | −0.103 | 0.700 | | | 0.264 | 0.000 |
| **Seven factors** | | | | | | | | | |
| | 0.628 | 0.280 | 0.036 | −0.174 | 0.199 | 0.084 | −0.103 | 0.935 | 0.000 |
| | 0.173 | 0.624 | −0.207 | −0.211 | −0.023 | 0.287 | 0.062 | 0.775 | 0.116 |
| | 0.201 | −0.123 | 0.671 | 0.203 | −0.284 | 0.041 | 0.016 | 0.775 | −0.116 |
| | −0.095 | −0.010 | 0.073 | 0.261 | 0.074 | 0.012 | −0.062 | 0.774 | 0.000 |
| | 0.120 | −0.252 | −0.195 | −0.046 | 0.585 | −0.389 | −0.358 | 0.255 | 0.000 |
| | −0.138 | 0.043 | 0.085 | −0.023 | 0.008 | 0.353 | −0.280 | 0.102 | 0.120 |
| | −0.287 | 0.182 | −0.164 | −0.046 | 0.035 | −0.229 | 0.597 | 0.102 | 0.120 |

| | Distribution Loadings $\boldsymbol{\Sigma}_\lambda$ | | | | | | | $\delta$ | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Five factors** | | | | | | | | | |
| | 0.034 | −0.054 | −0.025 | −0.015 | −0.072 | | | −0.038 | |
| | −0.054 | 0.231 | −0.015 | −0.036 | 0.202 | | | 0.105 | |
| | −0.025 | −0.015 | 0.065 | 0.080 | 0.043 | | | 0.028 | |
| | −0.015 | −0.037 | 0.080 | 0.145 | 0.040 | | | 0.023 | |
| | −0.072 | 0.202 | 0.043 | 0.040 | 0.235 | | | 0.125 | |
| **Seven factors** | | | | | | | | | |
| | 0.055 | 0.018 | 0.011 | 0.047 | −0.060 | −0.001 | 0.087 | 0.027 | |
| | 0.018 | 0.070 | 0.006 | 0.000 | −0.016 | 0.098 | 0.107 | 0.065 | |
| | 0.011 | 0.006 | 0.047 | 0.034 | −0.026 | −0.030 | 0.061 | 0.010 | |
| | 0.047 | 0.000 | 0.034 | 0.073 | −0.061 | −0.045 | 0.092 | 0.012 | |
| | −0.060 | −0.016 | −0.026 | −0.061 | 0.077 | 0.020 | −0.105 | −0.030 | |
| | −0.001 | 0.098 | −0.030 | −0.045 | 0.020 | 0.186 | 0.097 | 0.088 | |
| | 0.087 | 0.107 | 0.061 | 0.092 | −0.105 | 0.097 | 0.306 | 0.124 | |

The full sample of observations is used ($N = 144$ and $T = 200$), and we estimate the model parameters using the methods developed in section III. The columns indicated by eigen summarize the real and imaginary eigenvalues of the matrix $\boldsymbol{H}$ in decreasing order.

unbalanced, and the estimation methods of section III are adjusted accordingly, as discussed in the online appendix.

We consider the dynamic factor model (2) with stochastic loadings and with conditions of assumption 1. The loadings and factors are identified up to an orthonormal rotation matrix, which is sufficient for forecasting applications. We follow Stock and Watson (2012) and Jungbacker and Koopman (2015) and consider dynamic factor models with $r = 5$ and $r = 7$ factors. We first discuss the full sample parameter estimation results; then we discuss the details of our forecasting exercise and, finally, we present the results.

### A. Estimation Results

In table 3 we present the parameter estimates that define the the distributions of the loadings and the factors. The methods of section III are used to obtain the estimates. Prior to estimation, we have demeaned each time series. In the top panel of table 3, we present the estimates of the autoregressive matrix $\boldsymbol{H}$. From the eigenvalues, it follows that all the factors are estimated as stationary processes. The factors are persistent since the largest eigenvalues for both $r = 5$ and $r = 7$ are above 0.9. Additionally, we find some evidence for persistent cyclical behavior in the factors. For the five-factor model, one conjugate pair of complex eigenvalues is obtained with its real part equal to 0.809. The remaining two
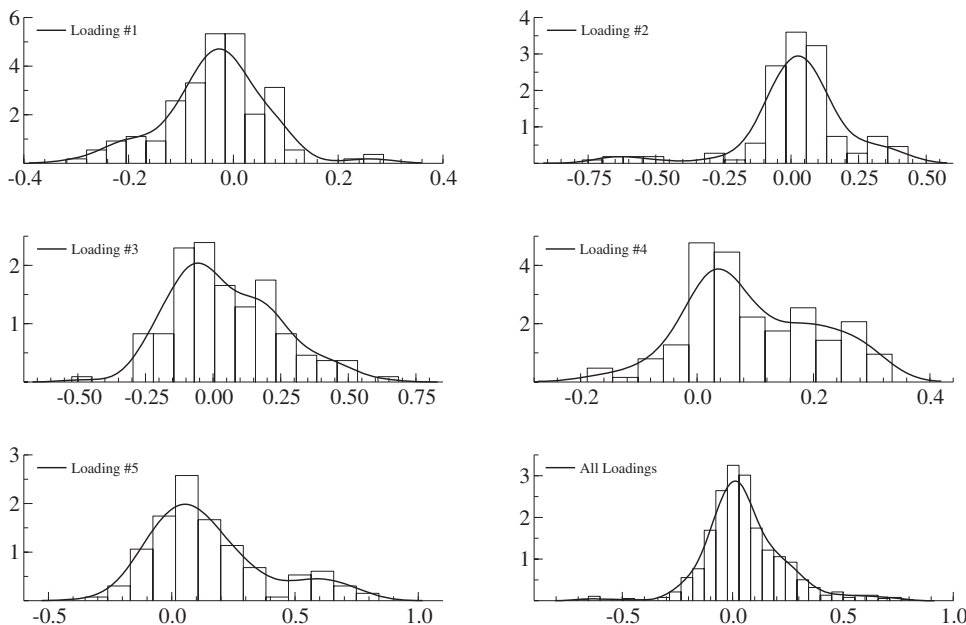
eigenvalues are relatively small. For the seven-factor model, we have two conjugate pairs with real parts equal to 0.775 and 0.102. Since the vector autoregressive (VAR) process of the factors imposes a 0 mean and an identity variance matrix, we can relate the individual coefficients in $\boldsymbol{H}$ to each other. However, as for any VAR analysis, individual coefficients in $\boldsymbol{H}$ do not have a clear interpretation.

In the bottom panel of table 3, we report the estimates for the means and variances of the loading matrices. The mean coefficients are small and typically close to 0. The diagonal elements of the variance matrices indicate that the variance in the loadings is small. However, the variances of the different loading columns are not equivalent, indicating that some of the factors are stronger relative to others.

In figure 1, we show the empirical density function and histograms of the estimates for the columns of the loading matrix. The third and fourth loading columns show multiple modes, whereas the first is slightly skewed to the left. All empirical distributions are quite similar to the normal distribution, with standard deviations between 0.1 and 0.2 approximately. However, the tails of the distributions are often heavier when compared to the normal distribution. In the lower right of figure 1, we show the empirical distribution for all loadings. This distribution is skewed to the right.

In the online appendix, we show the estimated factors and the correlations of the factors with the individual time series.

FIGURE 1.—EMPIRICAL DENSITY FUNCTION (SOLID LINE) AND HISTOGRAMS FOR THE POSTERIOR MODES OF THE COLUMNS OF THE LOADING MATRIX



We present the results per column and for all columns combined.

Since the model is identified subject to rotation, the empirical Bayes and maximum likelihood methods typically identify a different rotation. This implies that only the inner products are comparable. For a selection of time series we show the estimated inner products. We find that the estimates are close but not identical.

### B. Forecasting Study

The out-of-sample forecasting study for the panel of macroeconomic time series is designed as follows. For the models with $r = 5$ and $r = 7$ factors, we forecast each time series one, two, and four quarters ahead for 1985-1 to 2008-4. In total, we compute $m = 96$ out-of-sample predictions for each horizon and model. In particular, let the integer $n$ denote the sample split point (1984-4 for $h = 1$, 1984-3 for $h = 2$, and 1984-1 for $h = 4$). The forecasts are computed for $n+1, \ldots, n+m$ based on sequentially demeaned subsamples of the observations $y_1, \ldots, y_{n+j-h}$, for $j = 1, \ldots, m$. We estimate the parameter vector $\psi$ for each subsample using the methods of section III.

Based on the estimated parameter vectors, we compute the empirical Bayes posterior mode and maximum likelihood forecasts by

$$\hat{y}_{n+j} = \hat{\Lambda}\hat{\alpha}_{n+j} \quad \text{and} \quad \tilde{y}_{n+j} = \tilde{\Lambda}\tilde{\alpha}_{n+j},$$

where $\hat{\alpha}_{n+j} = \mathrm{E}(\alpha_{n+j}|y_1, \ldots, y_{n+j-h}; \hat{\Lambda}; \hat{\psi})$ is the posterior mode forecast for the factors and $\tilde{\alpha}_{n+j} = \mathrm{E}(\alpha_{n+j}|y_1, \ldots, y_{n+j-h}; \tilde{\Lambda}; \hat{\psi}_1)$ is the forecast for the factors based on the maximum likelihood estimates. These forecasts are computed for horizons $h = 1, 2, 4$. We add the sample

mean of the subsample to the forecasts before determining its accuracy.

As a measure of accuracy, we consider the mean squared error (MSE) of the out-of-sample forecasts. In particular we compute for each time series

$$\mathrm{MSE}_i^{\mathrm{PEB}} = m^{-1}\sum_{j=1}^{m}(y_{i,n+j} - \hat{y}_{i,n+j})^2 \quad \text{and}$$

$$\mathrm{MSE}_i^{\mathrm{MLE}} = m^{-1}\sum_{j=1}^{m}(y_{i,n+j} - \tilde{y}_{i,n+j})^2, \tag{18}$$

where $\hat{y}_{i,n+j}$ and $\tilde{y}_{i,n+j}$ are the $i$th elements of $\hat{y}_{n+j}$ and $\tilde{y}_{n+j}$, respectively. In this way, we compute mean squared error statistics for 144 time series for all forecasting horizons.

### C. Forecasting Results

In table 4 we present summary statistics for the relative MSE statistics, that is, $\mathrm{MSE}_i^{\mathrm{PEB}}/\mathrm{MSE}_i^{\mathrm{MLE}}$ for $i = 1, \ldots, 144$. For the model with $r = 5$ factors, the one-quarter-ahead forecasts of the empirical Bayes estimates are on average 8% more accurate; this decreases to 3% for the two-quarters-ahead forecasts and to 1% for the four-quarters-ahead forecasts. For the model with $r = 7$ factors, the relative gains are larger. In particular, for one-quarter ahead, the gain is 12%, which decreases to 5% and 2% for the two- and four-quarter-ahead forecasts, respectively. We have obtained some substantial forecasting gains for shorter forecast horizons. The relative improvement in accuracy of empirical Bayes forecasts declines as we forecast further into the future. This is not surprising since both the empirical Bayes and maximum likelihood estimates are based on

TABLE 4.—RELATIVE MEAN SQUARED ERROR STATISTICS FOR
OUT-OF-SAMPLE FORECASTING USING THE EMPIRICAL BAYES METHOD
AND THE MAXIMUM LIKELIHOOD METHOD

| | $r = 5$ Factors | | | $r = 7$ Factors | | |
|---|---|---|---|---|---|---|
| | $h = 1$ | $h = 2$ | $h = 4$ | $h = 1$ | $h = 2$ | $h = 4$ |
| **All series** | | | | | | |
| Mean quantiles | 0.912 | 0.979 | 0.994 | 0.874 | 0.951 | 0.978 |
| 0.05 | 0.762 | 0.828 | 0.861 | 0.696 | 0.766 | 0.819 |
| 0.25 | 0.893 | 0.930 | 0.940 | 0.847 | 0.915 | 0.931 |
| 0.50 | 0.979 | 0.984 | 0.992 | 0.932 | 0.993 | 0.993 |
| 0.75 | 1.028 | 1.025 | 1.024 | 1.023 | 1.026 | 1.033 |
| 0.95 | 1.176 | 1.123 | 1.142 | 1.129 | 1.102 | 1.187 |
| **Components (mean)** | | | | | | |
| GDP components | 0.901 | 0.925 | 0.932 | 0.832 | 0.966 | 0.950 |
| Industrial production | 0.958 | 0.973 | 0.981 | 0.838 | 0.894 | 0.911 |
| Employment | 0.829 | 0.829 | 0.876 | 0.838 | 0.857 | 0.866 |
| Unemployment rate | 0.988 | 0.969 | 0.975 | 1.031 | 0.994 | 0.989 |
| Housing | 0.853 | 0.943 | 0.992 | 0.727 | 0.844 | 0.940 |
| Inventories | 0.701 | 0.869 | 0.906 | 0.661 | 0.868 | 0.901 |
| Prices | 1.063 | 1.069 | 1.065 | 1.041 | 1.021 | 1.054 |
| Wages | 1.010 | 0.992 | 0.989 | 0.961 | 0.958 | 0.982 |
| Interest rates | 0.973 | 1.004 | 1.006 | 0.915 | 0.937 | 0.957 |
| Money | 1.302 | 0.997 | 1.001 | 1.458 | 1.004 | 1.006 |
| Exchange rates | 1.032 | 1.011 | 1.011 | 1.053 | 1.034 | 1.022 |
| Stock prices | 1.046 | 0.973 | 1.011 | 1.096 | 0.985 | 1.023 |
| Consumer expectations | 0.930 | 0.976 | 0.999 | 0.921 | 0.981 | 0.996 |

The results summarize the distribution of the statistics $\mathrm{MSE}_i^{\mathrm{PEB}}/\mathrm{MSE}_i^{\mathrm{MLE}}$, for $i = 1, \ldots, 144$ and forecast horizons $h = 1, 2, 4$.

the same vector autoregressive process for the factors. The out-of-sample forecasting variance for a larger forecast horizon is dominated by the contribution of the factors given that the loadings are time invariant.

We further present a selection of quantiles of the distributions. For the 0.05 quantile, we find gains between 30% and 15% depending on the forecast horizon. This indicates that for a modest number of time series, the gains are very large. On the other side of the distribution, the relative accuracy is somewhat more in favor of the maximum likelihood estimates. Here, we can even lose 18% for some time series.

Finally, we summarize the relative MSE statistics per category. We find that the largest gains are obtained for the real economic categories such as housing (this includes series related to housing starts), inventories, and employment. For the category of wages, prices, and money, we do not obtain improvements for the empirical Bayes methods. The largest gains are found for the time series that load strongly on the factors.

## VI. Conclusion

We have developed parametric empirical Bayes methods for the estimation of the dynamic factor model. The loadings, factors, and disturbances of the model are treated as latent stochastic variables, which follow gaussian distributions. For the estimation of the loadings and factors, we have developed a posterior mode algorithm that relies on standard methods for linear time series and regression models. The parameter vector is estimated by likelihood-based methods using a simple two-step implementation. The posterior means and

variances of the loadings are estimated by standard simulation methods. We circumvent the infinite variance problem by calculating the integral over the factors analytically. We emphasize that the computational effort for the empirical Bayes methods is only modestly larger when compared to standard maximum likelihood methods (see Doz et al., 2012, and Jungbacker & Koopman, 2015).

The methods are evaluated in a Monte Carlo study for dynamic factor models with different dimensions and different numbers of factors. For the common components, the empirical Bayes estimates always outperform the maximum likelihood estimates for the data-generating processes considered. The relative gains increase when the data-generating process includes irrelevant or weak factors. We have further illustrated our methods in an empirical application for forecasting macroeconomic time series. The empirical Bayes approach again dominates the maximum likelihood approach.

## REFERENCES

Aguilar, Omar, and Mike West, "Bayesian Dynamic Factor Models and Portfolio Allocation," *Journal of Business and Economic Statistics* 18 (2000), 338–357.

Anderson, Brian D. O., and John B. Moore, *Optimal Filtering* (Englewood Cliffs, NJ: Prentice Hall, 1979).

Bai, Jushan, and Kunpeng Li, "Statistical Analysis of Factor Models of High Dimension," *Annals of Statistics* 40 (2012), 436–465.

——— "Maximum Likelihood Estimation and Inference for Approximate Factor Models of High Dimension," this REVIEW 98 (2016), 298–309.

Bai, Jushan, and Serena Ng, "Determining the Number of Factors in Approximate Factor Models," *Econometrica* 70 (2002), 191–221.

——— "Large Dimensional Factor Analysis," *Foundations and Trends in Econometrics* 3 (2008), 89–163.

Banbura, Martha, and Michele Modugno, "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data," *Journal of Applied Econometrics* 29 (2014), 133–160.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz, "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics* 120 (2005), 387–422.

Besag, Julian. E., "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistical Society, Series B* 48 (1986), 259–302.

Bhattacharya, Anirban, and David B. Dunson, "Sparse Bayesian Infinite Factor Models," *Biometrika* 98 (2011), 291–306.

Bishop, Christopher M., *Pattern Recognition and Machine Learning* (New York: Springer, 2006).

Chan, J., and I. Jeliazkov, "Efficient Simulation and Integrated Likelihood Estimation in State Space Models," *International Journal of Mathematical Modelling and Numerical Optimisation* 1 (2009), 101–120.

Chan, J., R. Leon-Gonzalez, and R. W. Strachan, "Invariant Inference and Efficient Computation in the Static Factor Model," *Journal of the American Statistical Association* (forthcoming).

Doornik, Jurgen A., *Object-Oriented Matrix Programming Using Ox* (London: Timberlake Consultants Press, 2007).

Doucet, Arnaud N., Nando de Freitas, and Neil J. Gordon, *Sequential Monte Carlo Methods in Practice* (New York: Springer-Verlag, 2001).

Doz, Catherine, Domenico Giannone, and Lucrezia Reichlin, "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models," this REVIEW 94 (2012), 1014–1024.

Durbin, James, and Siem Jan Koopman, *Time Series Analysis by State Space Methods* (Oxford: Oxford University Press, 2012).

Efron, Brad, *Large Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction* (Cambridge: Cambridge University Press, 2010).

Efron, Brad, and Carl Morris, "Stein's Estimation Rule and Its Competitors," *Journal of the American Statistical Association* 68 (1973), 117–130.

Geweke, John F., "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica* 57 (1989), 1317–1339.

James, Willard, and Charles Stein, "Estimation with Quadratic Loss" (pp. 361–379), in J. Neyman, ed., *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley: University of California Press, 1961).

Julier, Simon J., and Jeffrey K. Uhlmann, "A New Extension of the Kalman Filter to Nonlinear Systems" (pp. 182–193), in I. Kadar, ed., *Signal Processing, Sensor Fusion, and Target Recognition VI* (Bellingham, WA: SPIE, 1997).

Jungbacker, Boris, and Siem Jan Koopman, "Monte Carlo Estimation for Nonlinear Non-Gaussian State Space Models," *Biometrika* 94 (2007), 827–839.

——— "Likelihood-Based Analysis for Dynamic Factor Analysis for Measurement and Forecasting," *Econometrics Journal* 18 (2015), C1–C21.

Kaufmann, Sylvia, and Christian Schumacher, "Bayesian Estimation of Sparse Dynamic Factor Model with Order-Independent Identification," Study Center Gerzensee working paper (2013).

Kim, H. H., and N. R. Swanson, "Forecasting Financial And Macroeconomic Variables Using Data Reduction Methods: New Empirical Evidence," *Journal of Econometrics* 178 (2014), 352–367.

Knox, Thomas, James H. Stock, and Mark W. Watson, "Empirical Bayes Regression with Many Regressors," Princeton University working paper (2004).

Koopman, Siem Jan, Andre Lucas, and Marcel Scharth, "Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State Space Models," *Journal of Business and Economic Statistics* 32 (2015), 114–127.

Koopman, Siem Jan, and Neil Shephard, "Exact Score for Time Series Models in State Space Form," *Biometrika* 79 (1992), 823–826.

Marron, J. Steve, and Matt P. Wand, "Exact Mean Integrated Squared Error," *Annals of Statistics* 20 (1992), 712–736.

McCausland, W. J., S. Miller, and D. Pelletier, "Simulation Smoothing for State-Space Models: A Computational Efficiency Analysis," *Computational Statistics and Data Analysis* 55 (2011), 199–212.

Meng, Xiao-Li, and Donald B. Rubin, "Maximum Likelihood Estimation via the ECM Algorithm: A General Framework," *Biometrika* 80 (1993), 267–278.

Mesters, Geert, and Siem Jan Koopman, "Generalized Dynamic Panel Data Models with Random Effects for Cross-Section and Time," *Journal of Econometrics* 154 (2014), 127–140.

Nakajima, Jouchi, and Mike West, "Bayesian Analysis of Latent Threshold Dynamic Models," *Journal of Business and Economic Statistics* 31 (2013), 151–164.

Newey, Whitney K., and Daniel L. McFadden, "Large Sample Estimation and Hypothesis Testing" (pp. 2111–2245), in R. F. Engle, and D. L. McFadden, eds., *Handbook of Econometrics* (Amsterdam: Elsevier, 1994).

Onatski, A., "Testing Hypotheses about the Number of Factors in Large Factor Models," *Econometrica* 77 (2009), 1447–1479.

——— "Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors," *Journal of Econometrics* 168 (2012), 244–258.

——— "Asymptotic Analysis of the Squared Estimation Error in Misspecified Factor Models," *Journal of Econometrics* 186 (2015), 388–406.

So, Mike K. P., "Posterior Mode Estimation for Nonlinear and Non-Gaussian State Space Models," *Statistica Sinica* 13 (2003), 255–274.

Stock, James H., and Mark W. Watson, "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association* 97 (2002a), 1167–1179.

——— "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics* 220 (2002b), 147–162.

——— "Implications of Dynamic Factor Models for VAR Analysis," (2005), unpublished manuscript.

——— "Dynamic Factor Models," in M. P. Clements, and D. F. Hendry, eds., *Oxford Handbook of Economic Forecasting* (Oxford: Oxford University Press, 2011).

——— "Generalized Shrinkage Methods for Forecasting Using Many Predictors," *Journal of Business and Economic Statistics* 30 (2012), 481–493.